Secure De-identification and Re-identification

William Landi PhD, R. Bharat Rao PhD

Computer Aided Diagnosis & Therapy Group, Siemens Medical Solutions, Malvern, PA

Keywords: HIPAA, Privacy, Security, Cryptography, De-identification, Re-identification

ABSTRACT

Today's healthcare organizations have both an ethical and legal responsibility for protecting patient privacy. However, the HIPAA privacy rule allows for the release of de-identified patient data for certain purposes. Secure encryption technology can be used to encrypt patient identified data so only the owners of the original data can re-identify the patient. It further allows consistent de-identification over episodic collection events.

1. INTRODUCTION

The HIPAA Privacy Rule [1] recognizes the need for healthcare organizations to share patient data. The term *de-identified data* refers to patient data from which all information that could reasonably be used to identify the patient has been removed (e.g., removing name, address, SSN, etc…). Our method de-identifies data so that only the owners of the original data and/or legally empowered entities can re-identify the patient.

With our method of de-identification, centers monitoring for natural or human induced disease outbreaks (e.g. a bioterrorism monitoring center) can analyze de-identified data while ensuring patient privacy under normal circumstances. Moreover, if an outbreak is detected the CDC or other legally empowered entities can immediately re-identify a patient by using their *private* key and then take the necessary action to confine the outbreak and to increase the chances of survival of the infected individuals.

Our method also supports the use of de-identified data in clinical trials and research studies. Our method ensures that even if the same patient is treated at different times, both episodes are recognized as being for the same patient (even when the data for each episode is collected at disparate times).

2. METHOD

Our method uses secure public key encryption technology (such as RSA[2]) to generate a public key based on one or more private keys (see Figure 1). Public keys are used for de-identification and are freely distributable with no privacy issues. The public keys are used to produce a unique reproducible encrypted version of sufficient patient identifying information to uniquely identify the patient.[1] We call this the *Encrypted ID*. The fact that the encryption is reproducible enables the same patient to be recognized at different collection times. Private keys must be kept secure as they are used for re-identification. It is possible for the government to maintain one master private key that could re-identify any de-identified data whereas a hospital could only re-identify data from its own patients.

The Encrypted ID is not human readable; e.g., one encryption of social security number 123-45-6789 results in:

> r:'<°M#ÿJÓ}ê?âsa¦T□óúR¿=h□†>b`□Á□FŠÌü±ß,þ?€½6ñHËbK°n
> {□□ýf9°÷?´×z«óDÛZlñfÙ1ftm□Ú?êr^GÉìñ□-□€Vöè³Nƒ□-H3*b
> ý0□îŒO¹h¸E□Ý¾Ã□ŒßX±□□ñ'

Thus, we introduce the concept of a *Study ID* which is a short string generated without any patient information. One method is to arbitrarily map the Encrypted ID to a unique integer (the Study ID). This mapping can be made publicly available without compromising patient privacy. The de-identification process involves replacing identified information with information based on the Study ID and storing the Encrypted ID. For example:

Identified Record	**De-identified Record**
Name: John Smith	Name: _23Name_
SS#: 123-45-6789	SS#: _23SS_
ICD9 Code: 482.4	ICD9 Code: 482.4
	EID: r:'<°M#ÿJÓ} … X±□□ñ'

Obviously, symmetric encryption technologies (same key to encode and decode) can be used in a similar way. For some applications symmetric key encryption would be sufficient, but in others the public/private key approach is superior. We have also extended our method to de-identify unstructured data, by replacing patient-specific strings in image headers and free text.

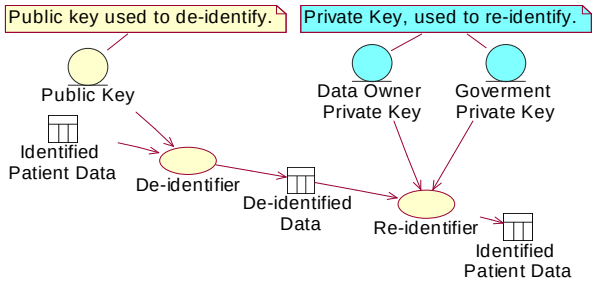

Figure 1: De-identification & Re-identification

REFERENCES

[1] HIPAA Privacy Rule. *45 CFR Parts 160 and 164 Standards for Privacy of Individually Identifiable Health Information; Final Rule.* Federal Register. Vol. 65, No. 250 (Dec. 2000), pp. 82462-82810.

[2] Rivest R.L., Shamir A., Adleman L.M. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. Communications of the ACM 21,2 (Feb. 1978), 120--126.

[1] When monitoring for disease outbreaks, additional information (e.g., address) may be included to support quick action.